Rethinking Artificial Intelligence creativity and ideation systems

Steve DiPaola
Simon Fraser University
Vancouver, BC
`sdipaola@sfu.ca`

Abstract

The visual and media arts community faces a significant challenge from the rise of generative visual AI, which poses a threat to their livelihoods. The current prompt-based generative AI (genAI) systems, such as Stable Diffusion, Midjourney, and DallE, are limited in their ability to provide sustainable financial opportunities for artists and designers, while also producing output that homogenizes unique cultural perspectives and styles. The importance of this topic lies in the cultural value that artists and designers bring to society, serving as a medium for expressing and preserving diverse traditions, beliefs, and values. We discuss a new approach to genAI and revised technological direction, as the current trajectory of AI development has the potential to displace human creators and devalue their skills.

1 Introduction

We propose techniques and have implemented a AI system to alter the current direction of genAI development and to create a pathway for creative professionals to utilize these systems to their advantage, both creatively and monetarily. The goal is to enhance (rather than constrain) the creative process of artists and designers by allowing them to ideate their own creative visions into a final artistic output of personal and/or societal significance and in the creative's own unique style. This toolkit also includes sustainable pathways for artists and designers to sell and/or license their work.

2 Rethinking Gen AI for more artist-centric creativity and ideation

We have a multi-year research in state-of-the-art in cognitive-based Deep AI techniques for creativity, and suggest these be applied to the creation of more artist-centric AI tools. This research goes beyond current efforts by utilizing our research in diffusion transformer and Large Language Modeling (LLM) systems, to create a system for artist-centric creative processes and workflows. Including:

1. Diffusion transformer based systems (Peebles, Xie 2023) and more aesthetic meaning centric datasets – where we use our cognitive modelling of creativity (DiPaola, 2012; DiPaola, McCaig, Gabora 2018; Utz, DiPaola 2020), aesthetic emotions (Abukhodair, ... DiPaola 2023) and bringing more human-centric meaning into datasets (and captions) (Salevati, DiPaola 2015; Yalçın, Utz, DiPaola, 2024)

2. Innovative new forms of LLMs and their fine tuning (He et al. 2023) to bring that deep meaning on some of these same areas: aesthetics, style, and emotion. Our new LLM system which we will continue improve, uses graph networks and new forms data augmentation systems rather than zero-shot generators. It uses robust cognitive and memory frameworks with data enrichment, episodic memory, and self-reflection (Gonzalez, DiPaola 2024).

Figure 1: An examples demonstrating movement through concept or style using our tool allowing for experimentation through complex associations along any axis in image-concept in latent space.

3. Better forms of local training of models such as LoRA and fine tuning (He et al 2023; Lin et al 2024; Choi, DiPaola, 2023) that can better encode an artistic style allowing new forms monetization and collaboration of an artist's body of work – where style and meaning are better built into the technologies.

Since their rise, visual genAI systems have been challenged repeatedly on a range of ethical issues such as their lack of transparency regarding their image training data and the associated copyright issues and the biases (incl. gender and racial) in their produced output. The often cited solutions for these ethical issues are improved development transparency, explainability of the applications and a reciprocal framework that considers the impacts on human creatives (Frosio, 2023).

The journey through AI's latent space offers a unique multimodal experience that reshapes artistic intentions and outcomes, questioning what is preserved, nurtured, lost, or irreversibly transformed in the interplay of the autographic (the artist's direct input) and the algorithmic (the AI's contribution). We explored these concepts in our cognitive-based creativity AI exploration system, and art analysis tools to parameterize a generative AI artistic painting process based on mood, style and emotion. Figure1 shows how portrait work brings in extra dimensions to this exploratory journey through a latent space, through how portraits give off strong emotional expression.

The implications of such technology on creative freedom and the essential relationship between the artist's subjective experiences and their artistic practices are profound. Our investigation utilizes a phenomenological approach to probe these emergent processes, highlighting how they mediate the expression of the mental image and expand the environment of creative practice beyond the intentional to include elements of serendipity and external perturbation (Sautoy, 2019). This paper builds upon foundational theories that view the artist and AI as co-creators, engaging in a process that is fundamentally improvisational and intersubjective. This approach mirrors the anticipatory and reflective modalities inherent in human artistic endeavor but extends them into the digital realm, where the non-linear, non-deterministic outputs of AI systems challenge and enrich the artist's creative agency. Moreover, the positioning of AI as a collaborator raises significant ethical questions about the nature of creativity and the role of the artist in the age of advanced technology.

We are now working with a prestigious design firm, FarmBoy Fine Arts (Farmboy) which ideates ideas for major hotels, hospitals and major car companies, as well as students from a well know art and design university, who with us are testing how they can explore new shared work of combined designer/artists through their style. See Fig1.

While our prototype tools are well underway with this evaluation work, we believe there are many unanswered questions in this very new way of creative ideation by cognitively journeying through a n-dimensional creative latent space, so are documenting questions and issues in this paper. Throughout this stage of the project, researchers need to work out the following aspects of this new creative system:

Figure 2: Moving through a latent metaphor space – here a "bird tree fly" hybrid landscape

1. What are the best new AI techniques and parameters for this type of LoRA (local training of a subset of creative work) training in style?

2. What image captioning techniques work best – captioning at training time or what is prompted at creation time? Do you use keywords or full design briefs to caption?

3. What size of dataset (input artworks) works best to maintain the style of an artist with a balance between style and variation for the designer to create?

4. How do these LoRAs work for the artist – can an artists create via their own personal dataset (or varied datasets/models) of their own work? How does this benefit individual artists?

5. How exactly does financial compensation work when a company contracts to use a LoRA of a specific affiliated artist/designer? (Farmboy already has experience with revenue sharing with their affiliated artists with existing plans that will be tested and modified as needed)

Current barriers impeding a solution include the limitations of prompt-based generative AI systems, which homogenize unique perspectives and styles, and the lack of sustainable financial opportunities for artists and designers within these systems. Current genAI systems like Midjourney and DalleE are based on using a text prompt that a user types into the system to generate an image or video. These prompts are not suited for many creative forms of expressions, as they are limited. As well aso are the original captions of the input images (or stories) in the dataset. These captions or keywords can be very superficial. Moreover, these text prompt systems do not allow artists to easily journey through a wide n-dimensional aesthetic space (the latent space of the dataset), which is necessary for creative exploration, to achieve the result they envision in their mind. In current systems, when the user changes the prompt due to dissatisfaction with current output – the output "jumps" around to another part of the search space (aesthetic latent space), preventing a slow and intentional exploration.

Our recommendations and prototype system allows for deeper meaning attributions that are linked to emotions, aesthetics, human philosophy, cultural history, and cognitive systems of art and vision (color palette, style, shapes, abstraction) for both stills and newer video to video techniques. Our existing system allows an artist to move from a current idea iteratively from image to image towards their final goal. Furthermore, it allows them to user label any part of their creative travel, where all directions of the journey (every choice) are saved and visual mapped out the journey, allowing the user to move forward or backwards through their creative journey.

Our system uses dynamic LoRAs, with a certain style - using the process to move through metaphor space – where "virtual cutout paper feather strokes" compose together to become, first birds, than trees and resolves by journey back in time and space over 100s of images to morph of a "bird tree fly" hybrid landscape, Fig2.

Preliminary work has shown that the labeled 3d mapping helps artists to better visualize their current creative journey. Our system can zoom in, and pan around latent space – creating a archived and live
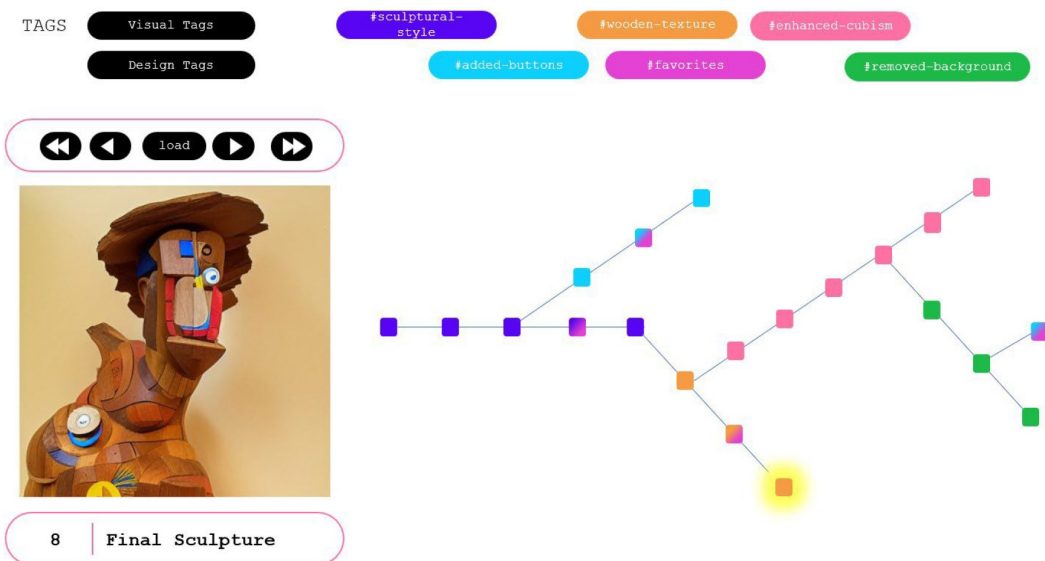
Figure 3: The mapping system shows forks that are taken (with tags and coloring) with in your creative decisions and allowing you to go back and forth – which allows you to better be free to experiment with ideation.
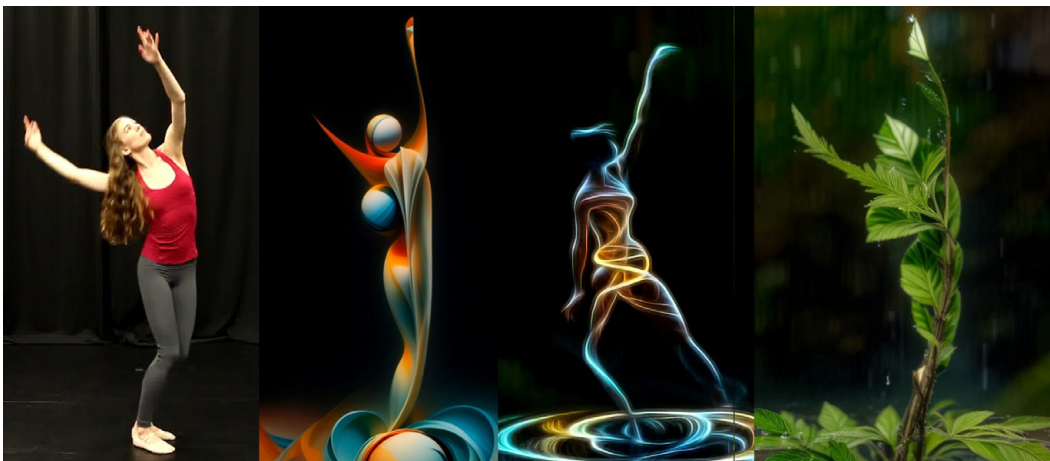

Figure 4: A dance sequence (video) can be transformed into different scenes - all where time-based aesthetic meaning is transformed.

creative journey. It then creates a dynamic map of every part of the journey. The mapping system shows forks that are taken (with tags and coloring) with in your creative decisions and allowing you to go back and forth – which allows you to better be free to experiment with ideation. Here you can click on any part of the map to bring up the user label (ie Final Sculpture) and output from any part of your paths as you journey. See Fig3. Our system also works for video to video sequences where meaning from say human movement (fig 4 left) can be understood and transferred to other time based forms but in their meaning space - for instance how a tree moves (fig 4 right) or as 3d forms (fig 4 middle). See Fig4.

Our novel strategy combines advanced AI diffusion systems with dynamic Lora style training, together with a artsit based front end user interface of intuitive labeling, mapping and exploring tools, all which helps to enhance the creative process and lower the cognitive load associated with complex generative AI. The goal of our evolving system (now in user testing) is for artist centric, open creative journeying and ideation, with a built in a mechanism to enable artists to license their work via novel latent space approaches for revenue-sharing.

4

References

1. Abukhodair N, Song M, DiPaola S. Designing a Wheel-based Assessment Tool to Measure Visual Aesthetic Emotions, Cognitive Systems Research, 23 pages, 2023.

2. Choi SK, DiPaola S. Art creation as an emergent multimodal journey in Artificial Intelligence latent space, ACM International Conference on Multimodal Interaction (ICMI), Workshop: Multimodal Affect and Aesthetic Experience, 7 pages, 2023.

3. DiPaola S, McCaig G, Gabora L. Informing Artificial Intelligence Generative Techniques using Cognitive Theories of Human Creativity. Procedia Computer Science, Special Issue: Biologically Inspired Cognitive Architectures, Vol 145, pp 158-168, 2018.

4. DiPaola, S. Computationally Modeling Creativity: Implementation and Response To A Computer Being Creative On Its Own, Proceedings of Conceptual Structure, Discourse and Language, 7 pages, 2012.

5. Frosio, G. Should We Ban Generative AI, Incentivise it or Make it a Medium for Inclusive Creativity? A Research Agenda for EU Copyright Law, 2023.

6. Gonzalez RA, DiPaola S. Exploring Augmentation and Cognitive Strategies for Synthetic Personae. ACM SigCHI, in Challenges and Opportunities of LLM-Based Synthetic Personae and Data in HCI Workshop, 9 pages, 2024.

7. He H, Cai J, Zhang J, Tao D, Zhuang B. Sensitivity-Aware Visual Parameter-Efficient Fine-Tuning. IEEE/CVF International Conference on Computer Vision (ICCV), 2023.

8. Lin L, Fan H, Zhang Z, Wang Y, Xu Y, Ling H. Tracking Meets LoRA: Faster Training, Larger Model, Stronger Performance, arXiv:2403.05231, 2024.

9. Peebles W, Xie S. Scalable diffusion models with transformers. InProceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4195-4205. 2023.

10. Salevati S, DiPaola S. A Creative Artificial Intelligence System to Investigate User Experience, Affect, Emotion and Creativity, Proc: Electronic Visualisation and the Arts, British Computer Society, 8 pages, 2015.

11. Sautoy, M. The Creativity Code: Art and Innovation in the Age of AI. Harvard University Press. 2019.

12. Yalçın OY, Utz V, and DiPaola S, Empathy through Aesthetics: Using AI Stylization for Visual Anonymization of Interview Videos, EmpathiCHI Workshop, ACM Conf on Human Factors in Computing Systems (CHI 24), 9 pages, 2024.

Supplemental Material

Our work in measuring cognitive based aesthetic emotional reception of artwork contributes to our main system:
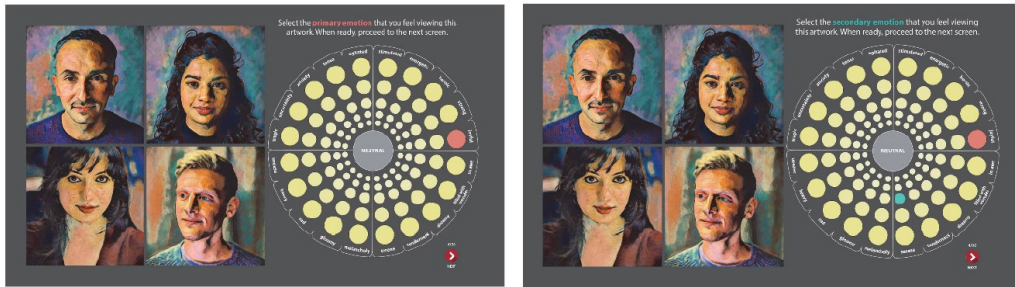

Our Visual Aesthetic Wheel of Emotion (VAWE) is a domain-specific device for measuring users visual aesthetic emotions to an artwork. The 20 aesthetic emotion terms are organized on a wheel-with spokes of the wheel representing the intensity users feel. Our systems allows for more natural, intuitive responses which differs from other instruments that require respondents to rate their feelings on a list of emotions terms as it organizes the terms to be rated on a theoretically justified two-dimensional system of valence and arousal.

One example of a portrait, with each of the 20 aesthetic emotional descriptors of VAWE. Here for one of our "self-portrait" participants.

Supplemental Material (continued):

Journeying or ideating through the systems can be in many forms, style, object, theme, emotion, ... as seen in these examples:
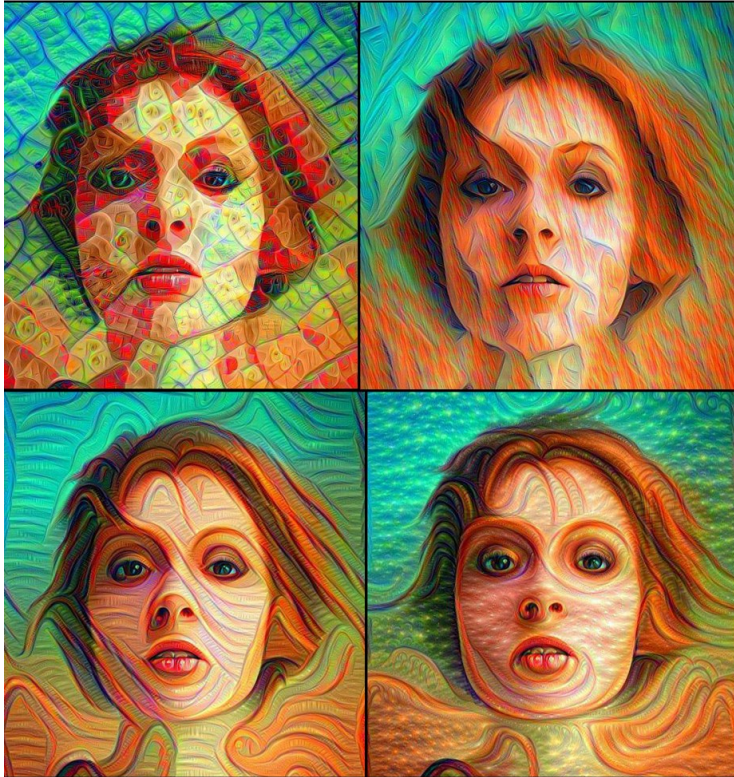
Here the same source portrait can move through different stylistic interpretations. Clockwise from upper left to lower right, they are: nature, fire, universe and water, all in the same ideation session with our system.

The ideation or journey through our system is not limited to simple style or color space, but can be any possible meaning or aesthetic level movement of ideas in the latent space (enhanced with LoRA setups). Here in the same session consisting of scores of output decisions, the artist can move through a "duo portrait space" with very different meanings. From "That couple, always together through the years" (left), to a painting entitled "Argument then fractured" (right).